DOI:10.11931/guihaia.gxzw201804001

菊叶香藜转录组数据库中 *FPPS* 基因的挖掘与生物信息学分析

付苏宏[1]，雷鸣[2]，张勇群[1*]，施静[1]，郝豆豆[1]

(1. 西藏自治区人民政府驻成都办事处医院 分子医学实验室，成都 610000；2. 西藏大学 理学院，拉萨 850000)

摘要：菊叶香藜精油具有潜在的药用价值，富含倍半萜类化合物，其生物合成分支点的关键酶为法尼基焦磷酸合酶(FPPS)。为了深入了解菊叶香藜法尼基焦磷酸合酶基因，本研究对菊叶香藜转录组数据库进行挖掘，获取了两条 *FPPS* 基因序列(*DsFPPS1* 和 *DsFPPS2*)，并对 *DsFPPS1* 和 *DsFPPS2* 编码蛋白的理化性质、结构、功能、系统进化进行了分析。结果表明，*DsFPPS1* 和 *DsFPPS2* 基因分别包含 1029 bp 和 969 bp 的开放阅读框，分别编码 342 个(DsFPPS1)和 322 个(DsFPPS2)氨基酸。DsFPPS1 和 DsFPPS2 位于线粒体内，未发现信号肽和跨膜结构，DsFPPS1 为稳定蛋白，DsFPPS2 为不稳定蛋白。氨基酸序列比对发现 DsFPPS1 和 DsFPPS2 序列相似性为 60.53%，均含有 5 个保守结构域和 2 个天冬氨酸富集区域。DsFPPS1 和 DsFPPS2 二级结构主要由α-螺旋构成，三级结构为由 8 个α-螺旋形成的α-螺旋束，但 DsFPPS2 的三级结构中缺少一个α-螺旋 A。系统进化树中 DsFPPS1 与藜科植物聚为一枝，与藜科植物遗传距离较近，而 DsFPPS2 单独聚为一枝。本研究通过对菊叶香藜转录组数据库中 FPPS 基因的挖掘与生物信息学分析，为菊叶香藜 FPPS 的功能研究及其倍半萜类化合物的生物合成研究奠定了一定的理论基础。

关键词：法尼基焦磷酸合酶，转录组，菊叶香藜，基因挖掘，生物信息学

Mining and Bioinformatic Analysis of *FPPS* Gene from *Dysphania schraderiana* Transcriptome Database

FU Suhong[1], LEI Ming[2], ZHANG Yongqun[1*], SHI Jing[1], HAO Doudou[1]

(1. *Molecular Medical Laboratory, Hospital of Chengdu Office of People's Government of Tibetan Autonomous Region*, Chengdu 610041, China; 2. *School of Science, Tibet University*,Lhasa 850000, China)

Abstract: *Dysphania schraderiana*, in the Chenopodiaceae family, is widely distributed in Lhasa (Tibet, China) and used as a traditional medicine. The essential oil of *Dysphania schraderiana*, which contains abundant sesquiterpenes compounds, appeared to possess potential medicinal value. Farnesyl pyrophosphate synthase (FPPS) is a key branch-point enzyme in biosynthesis of terpene. In order to reveal *D. schraderiana FPPS* gene, the transcriptome database of *D. schraderiana* was mined and two gene sequences (*DsFPPS1* and *DsFPPS2*) were obtained in this research. Subsequential protein physicochemical property, architectural feature, function and

基金项目：西藏自治区科技厅自然科学基金项目(2015ZR-13-5) [Supported by Natural Science Foundation of the Tibet Autonomous Region Department of science and technology (2015ZR-13-5)]。

作者简介：付苏宏(1993-)，女，四川内江人，硕士，科研助理，主要从事为分子生物学研究，(E-mail) fusuhong11@163.com。

***通讯作者** 张勇群，博士，副教授，主要从事藏药材基因与功能研究，(E-mail) yongqunzhang@yahoo.com。

phylogeny relationship analysis of DsFPPSs were also predicted and analyzed. The results indicated that *DsFPPS1* and *DsFPPS2* sequences contained an ORF span of 1029 bp and 969 bp respectively, encoding 342 (DsFPPS1) and 322 (DsFPPS2) amino acids respectively. The analysis of amino acid composition indicated that the dominant components of DsFPPS1 and DsFPPS2 are both nonpolar amino acids. The molecular weight of DsFPPS1 and DsFPPS2 were 39.68 kD and 36.76 kD, respectively. Isoelectric point were 5.11 and 5.65 for DsFPPS1 and DsFPPS2, respectively. Besides, DsFPPS1 protein was predicted to be a stable protein, but DsFPPS2 protein was predicted to be an unstable protein. The amino acid sequence analysis showed that DsFPPS1 and DsFPPS2 had no signal peptide and transmembrane region. The possible localization of DsFPPS1 and DsFPPS2 was both in mitochondria. DsFPPS1 and DsFPPS2 protein exhibited 60.53% sequence identity, and possessed five conserved domain (Ⅰ ~ Ⅴ) and two characteristic Asp-rich motifs (DDXXD). The amino acid sequence of DsFPPS1 has higher homology with *Chenopodium quinoa*, *Spinacia oleracea* and *Beta vulgaris* than DsFPPS2. In addition, the secondary structure of DsFPPS proteins mainly consisted of *α*-helixes, which resulted in a bundle of 8 *α*-helices in tertiary structure. However, tertiary structure analysis showed that DsFPPS2 protein was missing an *α*-helix compared to DsFPPS1 protein. The result of phylogenetic analysis indicated that phylogenetic relationships of DsFPPS1 protein were closer to Chenopodiaceae plants consistent with sequence alignment results, while DsFPPS2 protein was clustered alone in phylogenetic tree. In general, these results could lay a foundation for an insight to molecular function of DsFPPS and the synthetic biology of sesquiterpenes in *D. schraderiana*.

Key words: farnesyl pyrophosphate synthase, transcriptome, *Dysphania schraderiana*, gene mining, bioinformatics

菊叶香藜(*Dysphania schraderiana*)系藜科刺藜属植物，全草入药，具有平喘解痉、解毒、解表、止痛、止痒等功效，是我国传统的药用植物。现代研究发现菊叶香藜精油具有广泛的生物活性，本课题组在前期的研究中发现菊叶香藜精油对大肠杆菌和枯草芽孢杆菌的生长具有显著的抑制作用，并且呈现出一定的浓度依赖性；此外有研究发现菊叶香藜精油可浓度依赖性地抑制昆虫玉米象的活力(雷鸣, 2015)；现已有利用菊叶香藜精油治理植物螨虫的专利(刘志龙, 2015)。本课题组前期研究中利用 GC-MS 检测了菊叶香藜精油的化学成分进行了检测，除了精油中大量的脂肪族化合物(54.232 %)以外，倍半萜类化合物占有较大的比例(26.846%)。

倍半萜是一类包含有 15 个碳原子的天然化合物，即包含有 3 个异戊二烯单元，具有多种结构类型，是新药研发的主要来源之一(王佳, 2012)。植物萜类化合物主要是由细胞质中的甲羟戊酸(MVA)途径和质体中的 2-C-甲基-D-赤藓糖醇-4-磷酸(MEP)途径合成，这两条途径生成萜类化合物的共同前体异戊烯焦磷酸(isopentenyl pyrophosphate，IPP)和二甲基丙烯基二磷酸(dimethylallylpyrophosphate，DMAPP)，随后一分子的 IPP 和一分子的 DMAPP 结合生成香叶基二磷酸(geranyl diphosphate，GPP)，一分子的 GPP 和一分子的 IPP 在法尼基焦磷酸合酶(farnesyl diphosphate synthase, FPPS)的作用下生成倍半萜类化合物的前体物质法尼基焦磷酸(farnesyl diphosphate, FPP)，FPP 再在多种倍半萜合酶的作用下生成结构多样的倍半萜化合物(孙丽超, 2017)。

植物 FPPS 由基因家族编码，通常有多个 FPPS 亚型，例如拟南芥含有 *FPPS1* 和 *FPPS2* 两个基因，编码 3 个 FPPS 亚型：FPPS1S、FPPS1L 和 FPPS2，同时敲除两条 *FPPS* 基因会造成植物死亡，单个基因敲除不会对植物的生长发育造成影响，因此这种多基因模式可保证某一基因受损时植物仍能正常生存，但不同的亚型仍然具有不同的表达模式与活性，拟南芥

FPS1 存在于植物整个生命周期，FPS2 则主要在种子生长发育阶段发挥作用(Closa, 2010)，且 FPS2 具有更强的催化活性和热稳定性，对氯化钠的抑制作用也更敏感(Keim, 2012)。

FPPS 是萜类化合物生物合成分支点的关键酶，催化生成重要的中间产物 FPP 进而合成倍半萜类化合物。鉴于 FPPS 在倍半萜类化合物生物合成途径中的重要性，而目前尚未见菊叶香藜 *FPPS* 基因的研究报道，因此，为了深入探讨菊叶香藜的 FPPS，本研究对菊叶香藜花和叶转录组数据库进行挖掘以获取菊叶香藜 *FPPS* (*DsFPPS*)基因，并利用生物信息学的方法对该基因所编码蛋白的理化性质、亚细胞定位、结构特征、功能以及系统进化方面进行分析和预测，为今后深入研究该酶的结构和功能奠定理论基础。

1. 材料与方法

1.1 材料

菊叶香藜的两条 FPPS 基因来自于实验室前期菊叶香藜转录组研究，花组织转录组与叶组织转录组在 SRA (Sequence Read Archive) 数据库中的登录号分别为 SRX3145242 和 SRX3145241，其余物种相关的 FPPS 蛋白序列均来自 NCBI (National Center for Biotechnology Information) 数据库，物种及序列信息如下：藜麦 *Chenopodium quinoa* (CqFPPS: XP_021740746.1)、菠菜 *Spinacia oleracea* (SoFPPS: XP_021846881.1)、甜菜 *Beta vulgaris* (BvFPPS: XP_010675977.1)、野茶树 *Camellia sinensis* (CsFPPS: ANA11766.1)、杧果 *Mangifera indica* (MiFPPS: AFJ52720.1)、蓖麻 *Ricinus communis* (RcFPPS: AMN82836.1)、橡胶树 *Hevea brasiliensis* (HbFPPS: ANJ77846.1)、大戟 *Euphorbia pekinensis* (EpFPPS: ACN63187.1)、人参 *Panax ginseng* (PgFPPS: AAY87903.1)、竹节参 *Panax japonicas* (PjFPPS: AKN52395.1)、三七 *Panax notoginseng* (PnFPPS: AGS79228.1)、常春藤 *Hedera helix* (HhFPPS: APV45530.1)、西洋参 *Panax quinquefolius* (PqFPPS: ADJ68004.1)、薄荷 *Mentha x piperita* (MpFPPS: AAK63847.1)、撒尔维亚 *Salvia officinalis* Linn (SolFPPS: AQY54371.1)、丹参 *Salvia miltiorrhiza* (SmFPPS: ABV08819.1)、狭叶薰衣草 *Lavandula angustifolia* (LaFPPS: AGQ04160.1)、米团花 *Leucosceptrum canum* (LcFPPS: ALT07952.1)、黄芪 *Astragalus membranaceus* (AmFPPS: AID51444.1)、甘草 *Glycyrrhiza uralensis* (GuFPPS: ADE18770.1)、紫苜蓿(MsFPPS: ADC32809.1)、里氏木霉菌 *Trichoderma reesei* (TrFPPS: AFX82678.1)、新月金孢子菌 *Emmonsia crescens* (EcFPPS: POS87336.1)、地衣芽孢杆菌 *Emmonsia crescens* (BlFPPS: ARC74194.1)、荚膜红细菌 *Rhodobacter capsulatus* (RHcFPPS: ADE44162.1)、牛 *Bos Taurus* (BtFPPS: AAL58886.1)、黑猩猩 *Pan troglodytes* (PtFPPS: JAA28793.1)。

1.2 分析方法

1.2.1 蛋白基本性质分析

开放阅读框(ORF)通过 GenBank 的在线软件 ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/)进行查找。获取基因所编码的蛋白质氨基酸序列后，运用多种在线软件分析、预测蛋白的相关信息：蛋白的理化性质利用 ProtParam (http://web.expasy.org/protparam/)进行分析；蛋白的亲/疏水性分析应用在线软件 ProtScale (http://web.expasy.org/protscale/)；蛋白的信号肽分析应用在线工具 SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/)；蛋白的跨膜结构域预测应用在线程序 TMHMM (http://www.cbs.dtu.dk/services/TMHMM/)；蛋白的亚细胞定位预测运用在线软件 Cell-PLoc 2.0 (http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/)。

1.2.2 蛋白结构特征与功能分析

蛋白质的二级结构运用在线 SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html)软件进行预测；蛋白质的三级结构运用 Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index)在线软

件进行预测；利用在线数据库CDD (Conserved Domain Database) (https://www.ncbi.nlm.nih.gov/cdd/)进行保守位点分析。

1.2.3 蛋白氨基酸序列比对以及系统进化树构建

运用NCBI中的BLASTP (https://blast.ncbi.nlm.nih.gov/Blast.cgi)模块进行不同物种FPPS氨基酸序列的在线搜索，运用ClustalW进行多条FPPS氨基酸序列比对。采用邻接法(Neighbor-Joining)构建FPPS氨基酸序列的系统进化树，使用自展值(Bootstrap)重复检验1000次，系统进化树的构建和检验通过本地软件MEGA 7.0.14实施。

2. 结果与分析

2.1 菊叶香藜 *FPPS* 基因挖掘

通过对菊叶香藜转录组数据中核酸序列进行比对和功能注释，挖掘得到两条编号分别为c13053_g1和c12747_g1的*FPPS*候选基因，分别重命名为*DsFPPS1*和*DsFPPS2*。*DsFPPS1*长度为1 257 bp，其互补链转录翻译为FPPS蛋白，开放阅读框为1 029 bp，开放阅读框对应到*DsFPPS1*的位置为1 185~155 bp，编码342个氨基酸，编码蛋白命名为DsFPPS1；*DsFPPS2*长度为1 667 bp，包含一个969 bp的开放阅读框，位置从296~1 264 bp，编码322个氨基酸，编码蛋白命名为DsFPPS2(图1)。

A

```
        1         11        21        31        41        51        61        71
   1  TCTCCTTCGGGAGTAGAATAGATCATTCATCAATATAAGCTGTCTTTGCAGAAAATTAAATTACAAAAGCCACCTACATA

  81  ACACAACAAACCAGCTACAAAATAAATCAAGATGAATACAAAACCAATCAAAGAAGGCGAAAACACCAAGGACTTCACTT
                                                                              *  K  342

 161  CTGCCTCTTGTAAATTTTTGCCAAGAATGACTTCAGGACAGCTTGTACTGCTGTACTAGGATGTGCTTCAATAGATTCGA
       Q  R  K  Y  I  K  A  L  F  S  K  L  V  A  Q  V  A  T  S  P  H  A  E  I  S  E  I  315

 241  TCAATTTTTCATAACTCTTGCTCTCATAATCGGCAAACGCAGCCTGAAGATTGAGTACATTGTAGAGTTCTTTAACTCTT
        L  K  E  Y  S  K  S  E  Y  D  A  F  A  A  Q  L  N  L  V  N  Y  L  E  K  V  R   289

 321  GCGACACAAGCTGGGTCTTCTTTGCCGTAATTTTCATATAAGACTTGCTTTTGTTCATTGTTACAAAGCTCTAAGGCTTT
      A  V  C  A  P  D  E  K  G  Y  N  E  Y  L  V  Q  K  Q  E  N  N  C  L  E  L  A  K  262

 401  AACAACCAACCAAGAGCACTTAAAATCTTCAATATCGGTTCCAATCTTGCCAATAAATTCTGGGTCCCCAAAACAATCGA
       V  V  L  W  S  C  K  F  D  E  I  D  T  G  I  K  G  I  F  E  P  D  G  F  C  D  L  235

 481  GATAATCATCCTGTACTTGGAAATATATACCCATATCAATCAAAATGTTTTTGACATCGTCATGATCTTCAAGCTTCTCA
        Y  D  D  Q  V  Q  F  Y  I  G  M  D  I  L  I  N  K  V  D  D  H  D  E  L  K  E   209

 561  CCAGACATAAGTAAAGCACAAGCTACAGAAATATAAAATGAGTAATAAGCAGTTTTGTACTGGACAATGCGACGGTGCAA
      G  S  M  L  L  A  C  A  V  S  I  Y  F  S  Y  Y  A  T  K  Y  Q  V  I  R  R  H  L  182

 641  ATCCAAGGAATATTTTGAAAGATCTTTCTCTCCTTCGATGGTAGTGATCAGGTCAATCATCTGTCCACTGGCAGTCTGAA
       D  L  S  Y  K  S  L  D  K  E  G  E  I  T  T  I  L  D  I  M  Q  G  S  A  T  Q  F  155

 721  ATTCTACCTCATTAAACAAATCCAACAAGTCAACATAGTACTTCTTGCCCCTAAAATGCTTCCTGAGAATTCTAGGGATT
        E  V  E  N  F  L  D  L  L  D  V  Y  Y  K  K  G  R  F  H  K  R  L  I  R  P  I   129

 801  TGGTTTCGAAGTATTATACCATCGTTAACAGCAATCAAGCCAACCTTAGGAACTCTAAACCAGCAAGGCTGACCACGCCT
      Q  N  R  L  I  I  G  D  N  V  A  I  L  G  V  K  P  V  R  F  W  C  P  Q  G  R  R  102

 881  TGTATGAGAGTTATCCATGATGTCATCAAGAACAAGAAAATATGCTTGAAGCCATTCAATGCACCACCCCAGTGCAGAAG
       T  H  S  N  D  M  I  D  D  L  V  L  F  Y  A  Q  L  W  E  I  C  W  G  L  A  S  A  75

 961  CGAGAAAGATTTCATCATCAGTCAGCTCTTCCCCTTCCTTTAACAGCTTGTAACTGTCAATAACAGAAAGCCCACGATTT
        L  F  I  E  D  D  T  L  E  E  G  E  K  L  L  K  Y  S  D  I  V  S  L  G  R  N   49

1041  AGCTTCCCTCCAGGAACATTGTAATCCAGCATCCGGTCAACCCATTGACGAGAAGTGTGGGTCCATTCAAAAGCAGGGTC
      L  K  G  G  P  V  N  Y  D  L  M  R  D  V  W  Q  R  S  T  H  T  W  E  F  A  P  D   22

1121  ATTAAGAAGCTCAGATTTCAGAACTTGGTAAACTTCAAGGAACTTTGATCTCAGATCGCTCATTGTTTTTTGTTTTTGTT
       N  L  L  E  S  K  L  V  Q  Y  V  E  L  F  K  S  R  L  D  S  M                   1

1201  TTTATTTTTATTTTTGTTAAGAGATTGAAGAGAGTAAGAAGTGAGAGTATGGAGAAG
```

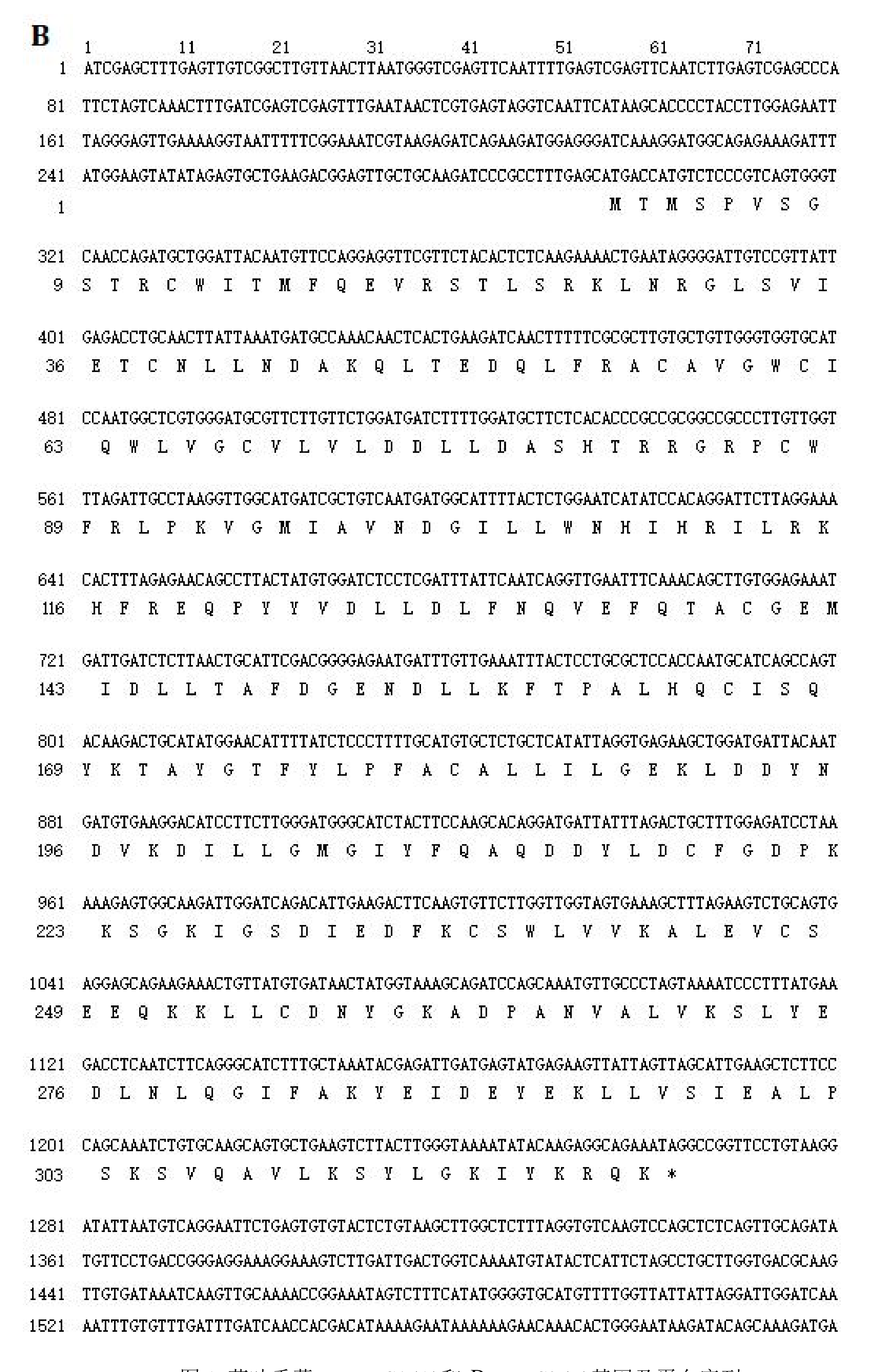


图 1 菊叶香藜 *DsFPPS1* (A)和 *DsFPPS2* (B)基因及蛋白序列

Fig. 1 The nucleotide sequence of *DsFPPS1* (A) and *DsFPPS2* (B) gene and their encoding

protein

2.2 DsFPPS 蛋白的理化性质分析

采用在线软件 ProtParam 分析 DsFPPS1 蛋白和 DsFPPS2 蛋白以及 3 种藜科植物的 FPPS 蛋白(藜麦 CqFPPS；菠菜 SoFPPS；甜菜 BvFPPS)的理化性质，结果如表 1 所示。DsFPPS1 蛋白与藜科植物 FPPS 蛋白较相似，氨基酸长度均为 342 aa；分子量在 39.67~39.74 之间；等电点(pI 值)在 5.01~5.12；酸性氨基酸比例在 15.5%~16.4%之间，碱性氨基酸比例在之间 13.8%~14.7%，极性中性氨基酸比例在 25.3%~26.5%之间，非极性氨基酸比例在 42.4%~44.4%之间，差异均较小；不稳定系数值均小于 40，属于稳定蛋白。而 DsFPPS2 蛋白与 DsFPPS1 蛋白以及藜科植物 FPPS 蛋白差异较大，氨基酸个数为 322 aa，长度短于其他 FPPS 蛋白；分子量为 36.76；等电点(pI 值)为 5.65；碱性氨基酸和极性中性氨基酸的比例与其他 FPPS 蛋白差异较小，比例分别为 13.1%和 27.3%，而酸性氨基酸的比例低于其他 FPPS 蛋白，为 13.1%，非极性氨基酸的比例高于其他 FPPS 蛋白，为 46.5%；不稳定系数值大于 40，为不稳定蛋白。

表 1 FPPS 蛋白理化性质

Table 1 The physicochemical properties of FPPS protein

蛋白名称 Name	氨基酸个数 Number of amino acids	分子量 (kDa) MW	等电点 pI	酸性氨基酸 (%) acidic amino acid	碱性氨基酸 (%) Alkaline amino acid	极性中性氨基酸 (%) Polar neutral amino acid	非极性氨基酸 (%) Nonpolar amino acid	不稳定系数 Instability index
DsFPPS1	342	39.68	5.11	15.5	13.8	26.3	44.4	36.71
DsFPPS2	322	36.76	5.65	13.1	13.1	27.3	46.5	41.58
CqFPPS	342	39.74	5.01	16.1	13.8	25.3	44.8	35.71
SoFPPS	342	39.74	5.12	16.4	14.7	26.5	42.4	28.58
BvFPPS	342	39.67	5.07	15.8	13.8	26.2	44.2	38.20

运用在线软件 ProtScale 分析 DsFPPS 蛋白的亲/疏水性，蛋白的每一位氨基酸的亲/疏水性用分值表示，正值则表示疏水，负值则表示亲水，其绝对值越大则其表示亲/疏水性越强。DsFPPS1 蛋白和 DsFPPS2 蛋白的亲/疏水性预测结果如图 2 示，在 DsFPPS1 蛋白氨基酸序列的第 202 位处有最高分值 2.611，表明其疏水性水性较强；第 101 位有最小分值-2.733，表明其亲水性较强；在 DsFPPS2 蛋白氨基酸序列的第 68 位处有最高分值 2.800，表明其疏水性水性较强；第 118 位有最小分值-2.578，表明其亲水性较强。

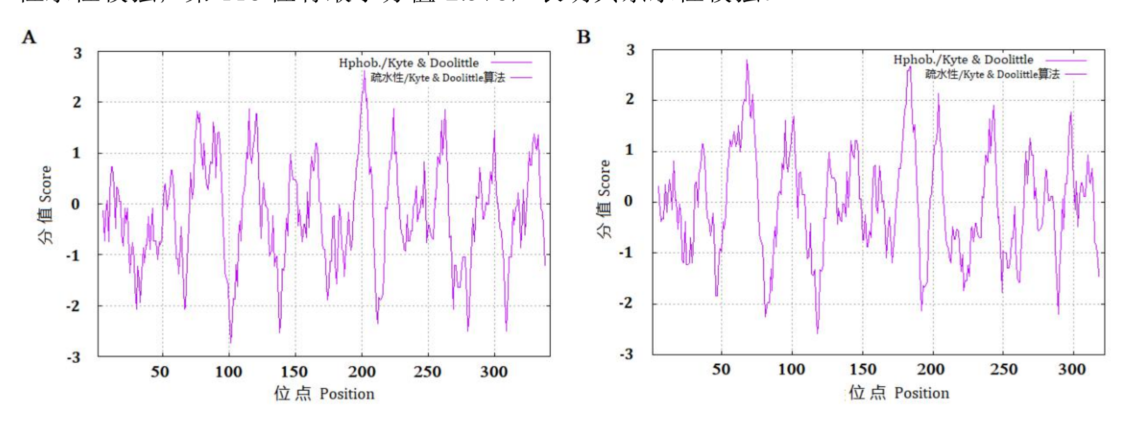


图 2 菊叶香藜 DsFPPS1 蛋白（A）和 DsFPPS2 蛋白（B）亲/疏水性分析

Fig. 2 Hydrophilic-hydrophobic property prediction of DsFPPS1 (A) and DsFPPS2 (B) protein

2.3 DsFPPS 蛋白的信号肽、跨膜结构域与亚细胞定位分析

信号肽是分泌蛋白 N 端所拥有的一段 15~30 个氨基酸残基的短肽，在分泌蛋白合成结束后将其切除，因此可以通过分析蛋白的氨基酸序列是否具有分泌肽来判断蛋白是否为分泌蛋白。利用在线软件 SignalP 4.1 对 *DsFPPS* 基因所编码的蛋白进行分析，结果显示 DsFPPS1 蛋白和 DsFPPS2 蛋白均不具有信号肽，属于非分泌蛋白(图 3A 和 3B)。然后，我们利用 TMHMM 在线软件分析 *DsFPPS* 基因编码蛋白的跨膜结构，结果显示 DsFPPS1 蛋白和 DsFPPS2 蛋白均不具有跨膜结构(图 4A 和 4B)。此外，Cell-PLoc 2.0 在线软件的亚细胞定位分析表明，DsFPPS1 蛋白和 DsFPPS2 蛋白均存在于细胞的线粒体内。

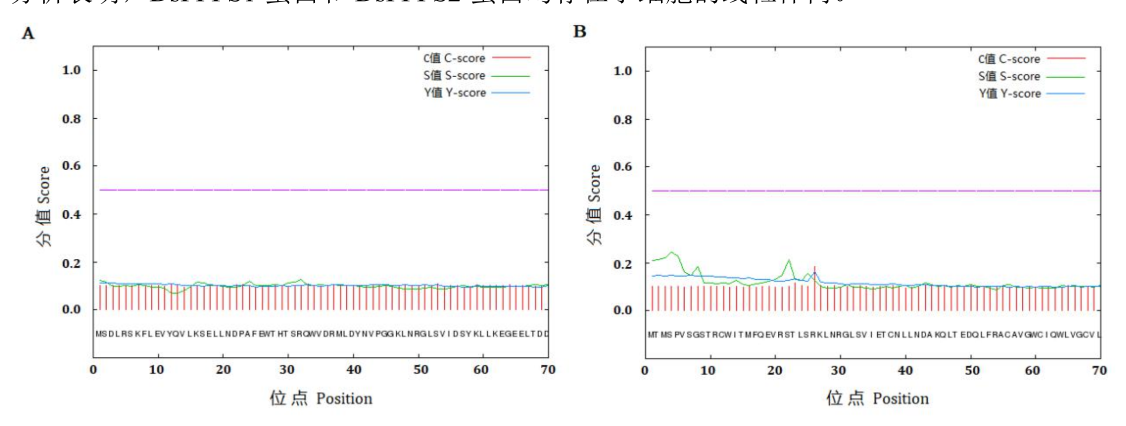


图 3 DsFPPS1 蛋白(A)和 DsFPPS2 蛋白(B)的信号肽分析

Fig. 3 Signal peptide prediction of DsFPPS1 (A) and DsFPPS2 (B) protein

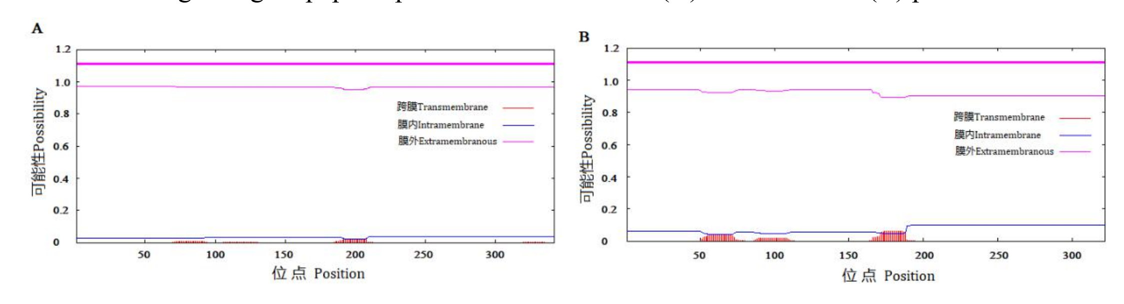


图 4 DsFPPS1 蛋白(A)和 DsFPPS2 蛋白(B)的跨膜结构域分析

Fig. 4 Transmembrane helix regions prediction of DsFPPS1 (A) and DsFPPS2 (B) protein

2.4 DsFPPS 蛋白的功能结构域分析

采用 NCBI 中的 CDD 工具、本地软件 ClustalW 以及相关文献(Chen, 1994; Guo, 2015; 李永波, 2012)寻找 FPPS 蛋白的保守结构域，结果显示 FPPS 家族蛋白共包含 5 个保守区域(Ⅰ~Ⅴ)，以及分别位于保守区域Ⅱ和Ⅴ的两个天冬氨酸富集区域(DDXXD，X 为任意氨基酸)：FARM (First Aspartic Rich Motif)和 SARM (Second Aspartic Rich Motif)，如图 5 所示。DsFPPS1 蛋白和 DsFPPS2 蛋白均属于 FPPS 家族，序列比对结果显示 DsFPPS1 蛋白和 DsFPPS2 蛋白具有 60.53%的相似性，均具有 FPPS 家族蛋白特征的两个天冬氨酸富集区域 FARM 和 SARM，以及序列保守区域Ⅰ~Ⅴ。天冬氨酸富集区域(DDXXD)是金属离子 Mg^{2+}结合位点，Mg^{2+}再与底物的磷酸基团结合，从而发挥催化作用(Christianson 2017)。DsFPPS1 蛋白对应的五个保守区域序列分别为 GGKLNR (45~50, Ⅰ)、EWLQAYFLVLDDIMDNSHTRRG (83~104，Ⅱ)、GQMIDL (160~165，Ⅲ)、KT (190~191，Ⅳ)、GIYFQVQDDYLDCFGDPEFIGKIGTDIEDFK (225~255，Ⅴ)，FARM 为 DDIMD (93~97)，SARM 为 DDYLD (232~236)。DsFPPS2 蛋白对应的五个保守区域分别为：SRKLNR (25~30，Ⅰ)、QWLVGCVLVLDDLLDASHTRRG (63~84，Ⅱ)、GEMIDL (140~145，Ⅲ)、KT (170~171，

Ⅳ)、 GIYFQAQDDYLDCFGDPKKSGKIGSDIEDFK (205~235， Ⅴ)， FARM 为 DDLLD (73~77)，SARM 为 DDYLD (212~216)。

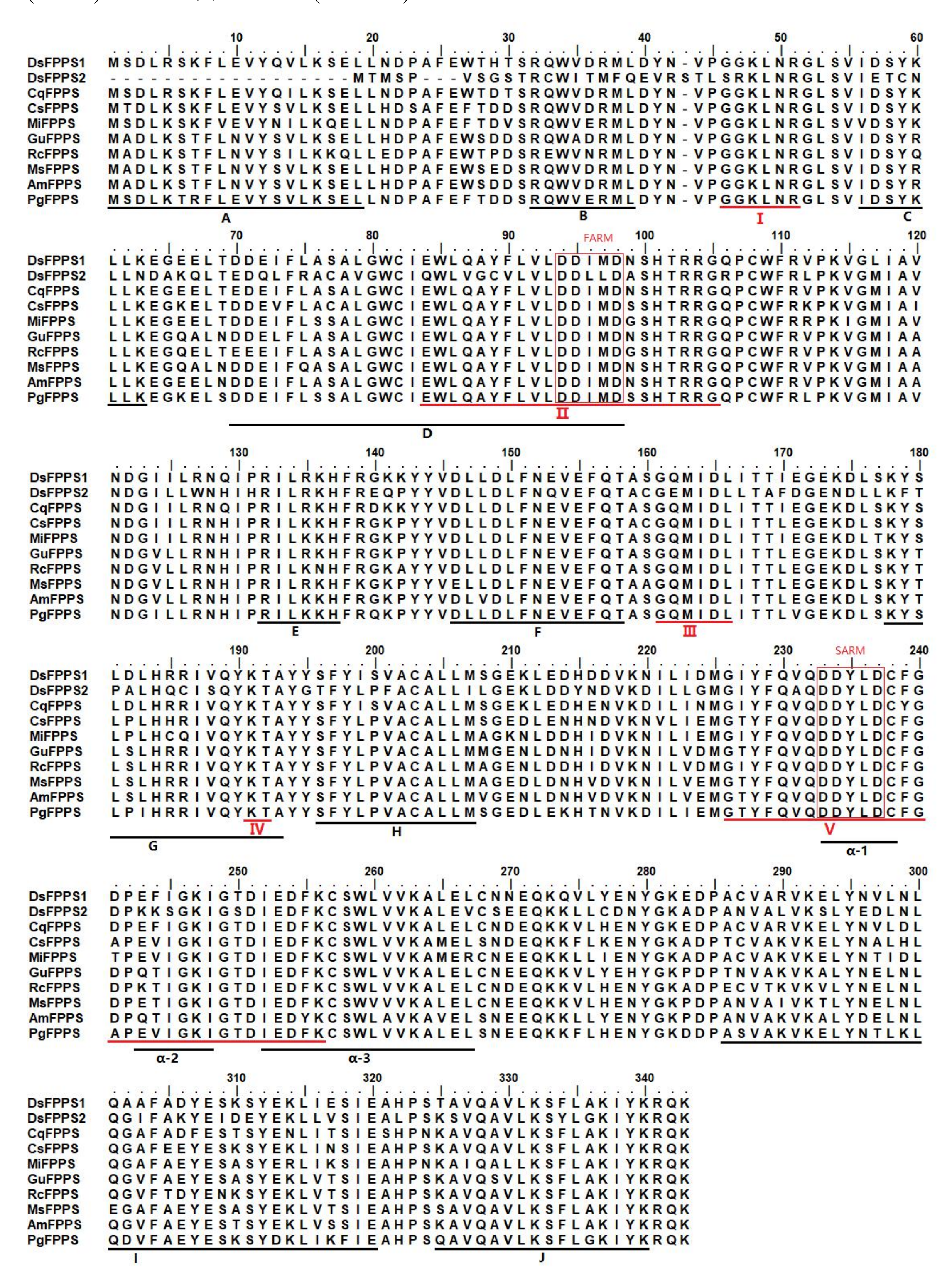


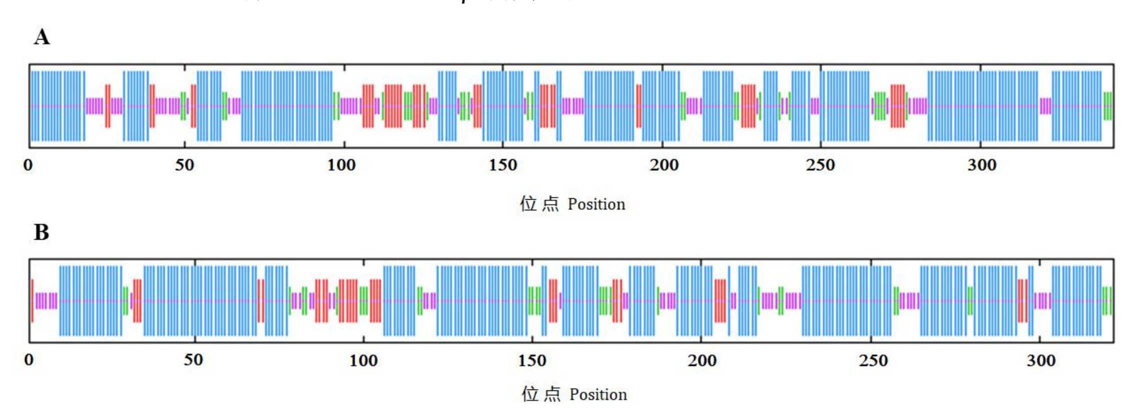
图 5 FPPS 蛋白多序列比对分析

Fig. 5 Multiple sequence alignment of FPPS protein

2.5 DsFPPS 蛋白的结构特征分析

使用 SOPMA 在线软件对 *DsFPPS* 基因所编码蛋白的二级结构进行分析和预测，结果表明(图 6) DsFPPS1 蛋白和 DsFPPS2 蛋白的二级结构组成类似，都是主要有α-螺旋组成，其比

例分别为 59.36%和 64.29%，此外，DsFPPS1 蛋白的二级结构还由 19.88%的无规则卷曲、11.70%的延伸链以及 9.06%的β-转角组成，DsFPPS2 蛋白的二级结构还由 15.53%的无规则卷曲、10.87%的延伸链以及 9.32%的β-转角组成。



注：蓝色为α-螺旋；红色为延伸链；绿色为β-转角；紫色为无规则卷曲。

Note: Blue lines represented α-helix; Red lines represented extended strand; Green lines represented β-turn; Purple lines represented random coil.

图 6 DsFPPS1 蛋白（A）和 DsFPPS2 蛋白（B）的二级结构分析

Fig. 6 Secondary structure prediction of DsFPPS1 (A) and DsFPPS2 (B) protein

蛋白质的空间构象是蛋白质发挥功能的重要结构，因此，我们利用在线软件 Phyre2 对 *DsFPPS* 基因所编码的 DsFPPS1 蛋白和 DsFPPS2 蛋白的三级结构进行预测，以 *Artemisia spiciformis* 的 FPPS 蛋白(PDB ID: 4KK2)为模板，采用线串法同源建模构建 DsFPPS1 蛋白(79%相似性)和 DsFPPS2 蛋白(63%相似性)的三级结构，结果如图 7 所示。与大多数的 FPPS 蛋白类似，DsFPPS1 蛋白的空间结构中总共包含有 13 个α-螺旋，其中有 10 个长α-螺旋(A~J)，α-螺旋 A 和α-螺旋 B 形成一个发夹结构，α-螺旋 C 至α-螺旋 J 共 8 个α-螺旋则形成α-螺旋束，围绕形成一个中央腔；此外还包含有 3 个小α-螺旋(α-1，α-2，α-3)，嵌入到α-螺旋 H 和α-螺旋 I 之间，酶的催化反应则发生在α-螺旋腔底。DsFPPS2 蛋白的三级结构与 DsFPPS1 蛋白类似，但只包含有 12 个α-螺旋，9 个长α-螺旋(B~J)和 3 个小α-螺旋(α-1，α-2，α-3)，不包含α-螺旋 A，因此缺少α-螺旋 A 和α-螺旋 B 形成的发夹结构。

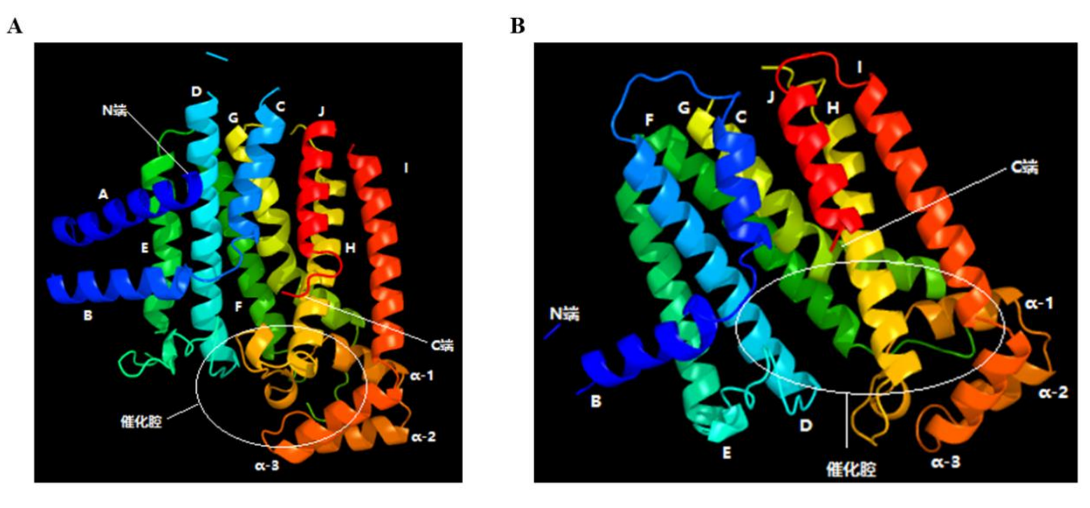


图 7 DsFPPS1 蛋白（A）和 DsFPPS2 蛋白（B）的三级结构模型预测

Fig. 7 Tertiary structure of DsFPPS1 (A) and DsFPPS2 (B) protein

2.6 DsFPPS 蛋白的系统进化分析

采用本地软件 MEGA 7.0.14 对 DsFPPS1 蛋白、DsFPPS2 蛋白和来自于 GenBank 数据库的 24 条 FPPS 蛋白序列进行分子系统进化分析，N-J 法(Neighbor-joining)构建系统进化树。结果表明(图 8)，FPPS 蛋白的系统进化树聚类分为植物、真菌、动物和细菌 4 个类群，具有较显著的种族特异性。细菌单独聚为一大枝，动物、真菌和植物聚为一大枝，该分枝中植物单独聚为一枝，另一枝又分化为动物和真菌两大枝。在植物分枝中，唇形科、五加科、藜科、蝴蝶花科、大戟科植物各自聚为一枝，菊叶香藜的 DsFPPS1 蛋白分布于藜科植物分枝中，说明其与藜科植物的亲缘关系较近，而菊叶香藜的 DsFPPS2 蛋白单独聚为一枝。

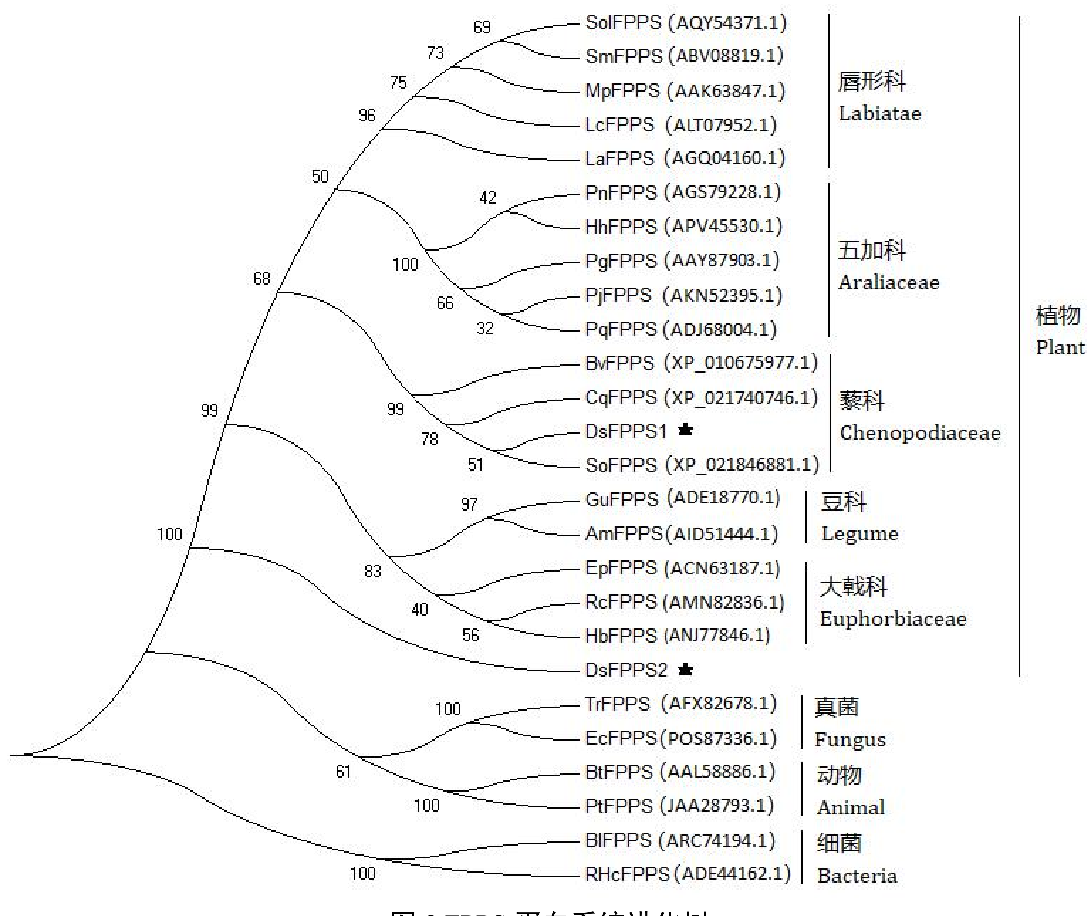


图 8 FPPS 蛋白系统进化树

Fig. 8 Phylogenetic tree of FPPS proteins

3. 结论与讨论

倍半萜类化合物作为植物的次级代谢产物具有较强的生物活性(朴英花, 2012)，例如青蒿素(Artemisinin, $C_{15}H_{22}O_5$)为倍半萜内酯药物，是有效的抗疟疾特效药(Tu 2011)；吉马酮(Germacrone, $C_{15}H_{22}O$)具有抗溃疡、抗肿瘤、抗炎、抗菌、止咳、利胆等多种功效(Wu, 2017)；β-桉叶醇(β-Eudesmol, $C_{15}H_{26}O$)可通过作用于瞬态受体电位锚蛋白 1 (TRPA1)和自主神经系统从而刺激食欲(Ohara, 2017)；雪松醇(Cedrol, $C_{15}H_{26}O$)可剂量依赖性地加速纤维母细胞生长以及增加 1 型胶原蛋白和弹性蛋白的生成(Mu, 2012)；缬草烯酸(Valerenic acid, $C_{15}H_{22}O_2$)可变构调节 GABA-A 受体从而发挥抗焦虑活性(Becker, 2014)。

FPPS 是植物萜类化合物生物合成途径中分支点上的关键酶，经 FPPS 的催化可生产倍半萜类化合物的前体，菊叶香藜精油中倍半萜类的生物合成需要此酶的参与。目前，随着转录组学研究的崛起，我们从包含有海量信息的转录组数据中挖掘 *FPPS* 候选基因，获得了两

条 *FPPS* 基因：*FPPS1* 和 *FPPS2*，并对这两条基因的开放阅读框进行了预测并进行了进一步的生物信息学分析，结果显示，FPPS1 蛋白和 FPPS2 蛋白序列相似度达 60.53%，均含有 FPPS 蛋白特征的 5 个保守结构域(Ⅰ保守)以及两个天冬氨酸富集区域(FARM 和 SARM)。然而，FPPS1 蛋白和 FPPS2 蛋白的稳定性存在差异，FPPS2 蛋白为不稳定蛋白，且 FPPS2 蛋白结构中缺少α-螺旋 A，因此其三级结构中缺少由α-螺旋 A 和α-螺旋 B 组成的发夹结构，该结构可能会影响其稳定性。在分子进化分析树的植物分枝中，FPPS2 蛋白与其他植物分开，单独聚为一枝，FPPS2 蛋白可能是一类新的 FPPS 亚型。植物 FPPS 通常由多基因家族编码，含有多个 FPPS 亚型，不同的亚型在植物不同的生长阶段和组织中发挥作用，拟南芥种子中 FPS2 的缺失可升高 HMGR(3-hydroxy-3-methylglutaryl CoA reductase，3-羟基-3-甲基戊二酰 CoA 还原酶)活性(Closa, 2010)，从而调节植物中“碳流”的流向，控制类异戊二烯种类的比例。

在前期的菊叶香藜转录组研究中发现，*FPPS1* 在花和叶组织中的表达均高于 *FPPS2*，且 FPPS1 蛋白的稳定性优于 FPPS2 蛋白，由此推测 FPPS1 蛋白广泛且长期存在于菊叶香藜植物中，而 FPPS2 的出现则具有“时-空”特异性，但菊叶香藜 FPPS 各亚型的功能、活性、表达以及调节机制仍需要更进一步的研究。总的来说，该研究对菊叶香藜转录组数据库中的 *FPPS* 基因进行了挖掘与生物信息学分析，为菊叶香藜 *FPPS* 基因今后的基础研究和开发应用奠定了理论基础。

参考文献

BECKER A, FELGENTREFF F, SCHRÖDER H, et al, 2014. The anxiolytic effects of a Valerian extract is based on valerenic acid [J]. BMC Complem Altern M, 14(1): 1-5.

CHEN A, KROON PA, POULTER CD, 1994. Isoprenyl diphosphate synthases: protein sequence comparisons, a phylogenetic tree, and predictions of secondary structure [J]. Protein Sci, 3(4): 600-607.

CHRISTIANSON DW, 2017. Structural and Chemical Biology of terpenoid cyclases [J]. Chem Rev, 117(17): 11570.

CLOSA M, VRANOVÁ E, BORTOLOTTI C, et al, 2010. The *Arabidopsis thaliana* FPP synthase isozymes have overlapping and specific functions in isoprenoid biosynthesis, and complete loss of FPP synthase activity causes early developmental arrest [J]. Plant J Cell Mol Biol, 63(3): 512.

GUO D, LI HL, PENG SQ, 2015. Structure conservation and differential expression of Farnesyl Diphosphate Synthase genes in Euphorbiaceous plants [J]. Int J Mol Sci, 16(9): 22402-22414.

KEIM V, MANZANO DA, DAVID, FERNÁNDEZ FJ, et al, 2012. Characterization of Arabidopsis FPS isozymes and *FPS* gene expression analysis provide insight into the biosynthesis of isoprenoid precursors in seeds [J]. PLoS ONE, 7(11): e49109.

LEI M, HE H, ZHANG PF, et al, 2015. Study on the extraction of *Chenopodium foetidum* essential oil and its inhibition against insect activity [J]. J Anhui Agric Sci, 43(28): 64-66. [雷鸣，何花，张鹏飞，等, 2015. 菊叶香藜精油的提取及对昆虫活力抑制的研究 [J]. 安徽农业科学, 43(28): 64-66.]

LI YB, FAN QQ, WANG BL, et al, 2012. Advances in the study of plant farnesyl pyrophosphate synthase (FPPS) gene [J]. Chin J Agr Biotechnol, 20(3): 321-330. [李永波，樊庆琦，王宝莲,等, 2012. 植物法呢基焦磷酸合酶基因(FPPS)研究进展 [J]. 农业生物技术学报, 20(3): 321-330.]

MU HJ, SUN GP, HWANG YL, et al, 2012. Cedrol enhances extracellular matrix production in Dermal Fibroblasts in a MAPK-Dependent manner [J]. Ann Dermatol, 24(1): 16-21.

OHARA K, FUKUDA T, ISHIDA Y, et al, 2017. β-Eudesmol, an oxygenized sesquiterpene, stimulates appetite via TRPA1 and the autonomic nervous system [J]. Sci Rep, 7(1): 15785-15081.

PIAO YH & PIAO HS, 2012. Research progress in biological activity of sesquiterpene compounds [J]. Occup

Health, 28(18): 2291-2293. [朴英花和朴惠顺, 2012. 倍半萜类化合物生物活性研究进展 [J]. 职业与健康, 28(18): 2291-2293.]

SUN LC, LI SY, WANG FZ, et al, 2017. Research progresses in the synthetic biology of terpenoids [J]. Biotechnol Bull, 33(1): 64-75. [孙丽超, 李淑英, 王凤忠, 等, 2017. 萜类化合物的合成生物学研究进展 [J]. 生物技术通报, 33(1): 64-75.]

TU YY, 2011. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine [J]. Nat Med, 17(10): 1217-1220.

WANG J, YOU S, ZHOU LN, 2012. Progress in research of sesquiterpenoids biotransformation [J]. J Shenyang Pharm Univ, 29(2): 76-84. [王佳, 游松, 周丽娜, 2012. 倍半萜生物转化的研究进展 [J]. 沈阳药科大学学报, 29 (2): 76-84.]

WU J, FENG Y, HAN C, et al, 2017. Germacrone derivatives: synthesis, biological activity, molecular docking studies and molecular dynamics simulations [J]. Oncotarget, 8(9): 15149-15158.

LIU ZL, LIU XC, SHI WP, et al, 2015-07-22. Essential oil of *Dysphania schraderiana* in the prevention and control of plant mites [P]. China. CN104782667A. [刘志龙, 刘昕超, 石旺鹏, 等, 2015-07-22. 菊叶香藜植物挥发油在防治植物螨中的用途 [P]. 中国发明专利. CN104782667A.]